Supporting Information

Development of Evolutionally Algorithm-Based Protein Redesign Method

Hiroki Ozawa[1, 3], Ibuki Unno[1, 3], Ryohei Sekine[1, 3], Sohei Ito[1] and Shogo Nakano[1, 2*]

[1]Graduate Division of Nutritional and Environmental Sciences, University of Shizuoka, 52-1 Yada, Suruga-ku, Shizuoka 422-8526, Japan

[2]PREST, Japan Science and Technology Agency, Saitama, Japan

[3]These authors contributed equally to this work

[*]Correspondence to Shogo Nakano (snakano@u-shizuoka-ken.ac.jp)

Keywords: Enzyme engineering, genetic algorithm, protein redesign, rational design

Contents

Material & Methods

Algorithm of GAOptimizer

The GAOptimizer software, developed using Python, incorporates the PyRosetta package for modifying and evaluating structural data[1]. To execute the software, two types of input data are required: a PDB file representing the template structure and library files containing homologous sequences of the template (Fig. 1). The sequence length of the template and the number of library files are denoted as *l* and *M*, respectively (Fig. 1). Upon completion of input data loading, the software employs a genetic algorithm to evolve the template structure, as illustrated in Fig. 1. Subsequently, a total of *N* elite structures are generated through the following procedures, where *N*, defined by the user, represents the generation number. Here, we defined n^{th} generation as $[1, N] \in n$.

At $n = 1$, the template structure serves as the sole parent structure ($P = 1$), and a total of 100 child structures are generated through continuous execution of Procedure 1 and random mutation (Fig. 1). In Procedure 1, an appearance frequency matrix for 20 L-amino acids (AAs) is created using the following steps: 1) The PDB data (*p* in Fig. 1, $[1, P] \in p$) corresponding to the template structure at $n = 1$ is converted to sequence data using PyRosetta and combined with randomly selected sequences from the library (*m* in Fig. 1, $[1, M] \in m$); 2) The combined sequences are aligned using multiple sequence alignment (MSA) with the MAFFT software [2], and the results are converted into an appearance frequency matrix of amino acids (Fig. 1). The matrix consists of $20 \times l$ elements, with rows representing residue numbers of the template structure denoted as *p* and columns representing AA frequencies calculated from the MSA results. The frequencies are represented by the first letter of each amino acid in alphabetical order; for example, the frequencies for A, C, D, and Y are represented in the 1^{st}, 2^{nd}, 3^{rd}, and 20^{th} columns, respectively. In the random mutation process, the software introduces consensus mutations into the parent, with the number of mutations randomly determined within the range of 0 to int($l \times 0.015$). The software first randomly selects residues on the template where mutations will be introduced. Consensus residues are identified by analyzing the matrix row at the site; in this study, residues with a frequency exceeding 30% (threshold value) are considered consensus residues. If multiple candidates for consensus residues are present, the software randomly selects one of the mutations from among them. The random mutations were introduced if the threshold value is set to 0%.

Upon generating a total of 100 child structures, the software calculates scores for each structure (Fig. 1). In this study, we employed two scores as fitness functions: the Rosetta energy unit (REU), calculated by PyRosetta, and the HiSol score, as depicted in Fig. 4. The HiSol score is calculated based on the equation presented in Fig. 4B, and a penalty score is assigned to child structures with REU values higher than that of the template structure.

Structures with the best scores are selected and saved as the elite structures of the nth generation. The subsequent step involves preparing a new population for the $(n+1)^{th}$ generation from the child structures. The selection process employs tournament selection, wherein: 1) an elite structure is chosen from five randomly selected child structures; 2) the selection process is repeated 30 times. For the selection, sampling with replacement is adopted. The total of 30 selected structures will serve as parents for the $(n+1)^{th}$ generation ($n = n+1$ and $P = 30$ in Fig. 1).

Under the condition of $n > 1$, child structures are generated through two distinct mutation processes (Fig. 1), with a total of 30 parent structures ($P = 30$) present (Fig. 1). One process involves random mutation, similar to the approach employed at n = 1. The random mutation process is adopted with the following probabilities: 100 - (n-1) × 2% for $n < 35$ and 30% for $n \geq 35$. The other process entails recombination (Fig. 1). In this process, after determining two crossover points, the p_1 structure is mutated to adopt the residues of the p_2 structure between the two points (red squared region in Fig. 1); p_1 and p_2 structures are randomly selected from the parents ([1, P] ∈ p_1 and p_2 in Fig. 1). The mutated p_1 structure becomes a child structure. The recombination mutation process is performed with the following probabilities: (n-1) × 2% for $n < 35$ and 70% for $n \geq 35$. Upon completing the generation of child structures, the same scoring and selection processes used at $n = 1$ are applied. The evolution process is repeated until the nth value reaches N. GAOptimizer can be accessed at the following website: ().

Design of the artificial ALPs by PROSS and GAOptimizer

Homolog sequences of SCAPase (PDB ID: 3A52), which is a psychrophilic alkaline phosphatase (ALP) from *Shewanella* sp.[3], were obtained by Blastp analysis; the E-value and database were set to $1.0E^{-3}$ and non-redundant, respectively. A total of 166 sequences (called as the curated-lib) were obtained by omitting sequences that have < 50% identity and more than 5% longer or shorter than the sequence of SCAPase (Table S1). Artificial ALPs were designed using two software: GAOptimizer and PROSS (Table S2). The GAOptimizer was run with following conditions: $M = 1$ (library was the curated-lib) and $N = 1{,}000$. The energy-minimized structure of 3A52 (3A52-EM.pdb) was used as the template (Table S2). REU score was adopted as the fitness function. Two designed ALPs were obtained after the calculation: GaALP-26gen and GaALP-1000gen which are elite structures at 26^{th} and 1000^{th} generation, respectively (Fig. S1 and Table S3). The PROSS software (https://pross.weizmann.ac.il/step/pross-terms/)[4, 5] was used to design two artificial ALPs (PROSS-2 and PROSS-3) by submitting the 3A52-EM.pdb and the curated-lib as input data (as shown in Table S2). The number of mutations was similar between GaALP-26gen and PROSS-2, or between GaALP-1000gen and PROSS-3 (as shown in Table S2), respectively.

As a reference, artificial ALP, called PROSS-1, was generated with PROSS by submitting 3A52.pdb as input. The library was automatically selected by the software. The sequence identity among the SCAPase and the designed ALPs (Table S3) was shown in Table S4.

Overexpression and purification of the ALPs

A gene encoding five artificial ALPs (Table S3) was synthesized by ordering to GenScript and subcloned into pET22b vector at *Nde*I and *Xho*I sites. The genes were optimized for the *E.coli* expression system. The plasmids transformed into BL21(DE3) strain. The strain was cultivated in 1L liquid LB medium containing 30 μg/mL carbenicillin at 37 °C. The temperature was lowering to 23 °C when O.D.600 value reached to the range of 0.6–0.8; the IPTG was added into the liquid medium at the final concentration of 0.5 mM. After continuing the cultivation for 16 hr, the cells were harvested by centrifugation. The cells were suspended in bufferA-ALP (25 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.25 mM $MgCl_2$, 0.025 mM $ZnSO_4$, and 5% (v/v) glycerol). After the suspended cells were sonicated, the supernatant was collected by centrifugation at 11,000 g for 30 min. The supernatant was applied to a HisTrap-HP (Cytiva, MA, USA) column which was already equilibrated by bufferA-ALP. After the column was washed with bufferA-ALP containing 50 mM imidazole, the samples containing ALPs were eluted with bufferA-ALP containing 400 mM imidazole. The eluted samples were applied to Superdex 200 Increase 10/300 GL column (Cytiva, MA, USA) equilibrated by bufferA-ALP. The purity of ALPs was confirmed by SDS-PAGE. The concentration of ALPs was estimated by measuring absorption at 280 nm by NanoDrop One (Thermofisher Scientific, USA). Protein expression level of ALPs was estimated by measuring the amount of the ALPs obtained after the purification by Ni-affinity chromatography (Table S5).

Analysis of enzymatic properties of the five artificial ALPs

The enzymatic activity of the five artificial ALPs was measured by quantifying the concentration of the produced *p*-nitrophenol (pNP) with color-developed method. All of the following experiments were performed in three replicates. The following assay buffer was prepared: 0.1 M diethanolamine (DEA) (pH 9.8), 10 mM *p*-nitrophenyl phosphate, 10 μM $ZnCl_2$, 0.5 mM $MgCl_2$. The assay buffer was incubated at 25 °C prior to measure the activity. The reaction initiated by adding the purified ALP solution into the assay buffer dispensed to cuvette. The initial velocity of the ALPs was measured by monitoring time-dependent absorption change at 405 nm derived from pNP by UV-visible spectrometer (UV-2450, Shimadzu). The molar extinction coefficient of pNP at 405 nm was as follows: ε_{405} = 18,500 $M^{-1}*cm^{-1}$.

The kinetic parameters of ALPs were estimated by utilizing the following assay buffer: 0.1 M diethanolamine (DEA) (pH 9.8), 0.01–10 mM *p*-nitrophenyl phosphate, 10 μM $ZnCl_2$, 0.5 mM $MgCl_2$. The enzyme kinetic parameters were estimated by fitting the obtained initial velocity to the Michaelis-Menten equation with the non-linear least square method by ORIGIN software (Fig. S2A). The calculated parameters were represented in Table S6.

Thermal stability of the artificial ALPs was analyzed by PEAQ-DSC (Microcal. Inc.). 1.2–9.2 μM for the ALPs in the bufferA-ALP were utilized as the samples. The samples were heated ranging from 60–100 °C that the increase of temperature speed was set to 90 °C per hour (Fig. S2B). The T_m values were estimated by fitting the obtained data (Table S6).

Design of the artificial TDHs by GAOptimizer

Two artificial L-threonine 3-dehydrogenases (TDHs) were designed with GAOptimizer as following procedures. The input data were firstly prepared: structure of TDH from *Cupriavidus necator* (CnTDH, PDB ID: 3WMX) which was energy-minimized by PyRosetta and five of the curated libraries prepared in previous study [6] were adopted as the template structure and the libraries (*M* = 5), respectively. Utilizing the input data, CnTDH was computationally evolved by GAOptimizer through 2,000 generation (N = 2,000); REU was adopted as the fitness function. In the elite structures, TDHs at 26th (GaTDH-26gen) and 2000th generations (GaTDH-2000gen) were adopted as research targets, respectively (Fig. S5B). We also designed another GaTDH by changing the library to only “MLYM library” (Fig. S5A). Although the REU values were decreased during the evolutional process (Fig. S5A), the TDH at 500th generation was expressed as insoluble form as well as the full consensus TDH designed by using the MLYM library. The sequence identity among the CnTDH and the designed TDHs (Table S7) were shown in Table S8.

Overexpression and purification of the TDHs

A gene encoding the two GaTDHs (GaTDH-26gen and GaTDH-2000gen, Table S7) was synthesized by ordering GenScript and subcloned into pET15b vector at *Nde*I and *Bam*HI restriction site. The genes were optimized for the expression by *E.coli* system. The plasmids coding CnTDH and FcTDH were obtained from the previous study; here, the FcTDH is corresponding with FcTDH-VLYM which had the best enzymatic properties in the designed TDHs [6, 7]. The plasmids transformed into BL21(DE3) strain. The strain was cultivated in 1L LB browth containing 30 μg/mL carbenicillin at 37 °C. The flask containing the strain was kept on ice for 30 min when the O.D.600 value was reached to the range from 0.6 to 0.8. After addition of 0.5 mM IPTG into the browth at final concentration, the cultivation was continued for 16 hr at 23 °C. The cells were collected by centrifugation and suspended into the bufferA-

TDH (10 mM potassium phosphate (pH 7.0) and 50 mM NaCl). After sonication of the suspended cells was completed, the supernatant was collected by centrifugation at 11000 g for 30 min. The supernatant was applied to HisTrap-HP column which was equilibrated by bufferA-TDH. After the column was washed with bufferA-TDH containing 50 (for GaTDH-2000gen) or 70 mM imidazole (for GaTDH-26gen and FcTDH), the samples containing the TDHs were eluted by bufferA-TDH containing 500 mM imidazole. The eluted samples were concentrated to 500 μL utilizing centrifugal filter unit, Amicon Ultra-15 (Merck Millipore). The concentrated samples were applied to gel filtration column (Superdex 200 Increase 10/300 GL, Cytiva, MA, USA) which was equilibrated by bufferA-TDH. The purity of the TDHs was confirmed by SDS-PAGE, and the concentration was estimated by measuring absorption at 280 nm by NanoDrop One (Thermofisher Scientific, USA). Protein expression level of TDHs was estimated by quantifying the amount of the TDHs obtained after purification by Ni-affinity chromatography (Table S9).

Analysis of enzymatic properties of the TDHs

The enzymatic activity of TDHs was measured by quantifying the concentration of NADH with color-developed method. All of the following experiments were performed in three replicates. The contents of the assay buffer were as follows: 0.1 M buffer, 2.5 mM NAD^+, and 20 mM L-Thr. The following buffer was used to estimate optimal pH value of the TDHs: Bis-tris HCl (pH 6.0, 6.5 and 7.0), Tris-HCl (pH 7.0, 7.5, 8.0, 8.5, 9.0 and 9.5), Glycine-KOH buffer (pH 10.0, 10.5 and 11.0), respectively (Fig. S6A). The assay buffer was incubated for 30 min at 30 °C condition. The reaction was started by adding the purified TDHs samples into the assay buffer which was dispensed into the cuvette. The initial velocity of the TDHs was measured by monitoring time-dependent absorption change at 340 nm derived from NADH by UV-visible spectrometer (UV-2450, Shimadzu). The molar extinction coefficient of NADH at 340 nm was as follows: ε_{340} = 6,300 $M^{-1}*cm^{-1}$.

The kinetic parameters of TDHs were estimated by utilizing the following assay buffer: 0.1 M buffer (Tris-HCl [pH 8.5] buffer for GaTDH-26gen and Tris-HCl [pH 9.0] buffer for GaTDH-2000gen, respectively), 2.5 mM NAD^+, and 0–20 mM L-Thr. The enzyme kinetic parameters were estimated by fitting the initial velocity to the Michaelis-Menten equation with non-linear least square method by ORIGIN software (Fig. S6B). The parameters were represented in Table S10.

Thermal stability of the four TDHs (CnTDH, FcTDH, GaTDH-26gen and GaTDH-2000gen) was analyzed by PEAQ-DSC (Microcal. Inc.). Prior to start the measurement, the TDH samples were purified by gel filtration chromatography which was equilibrated by bufferB-TDH (10 mM potassium phosphate [pH 7.0], and 150 mM NaCl). 140–250 μM of the TDHs

in bufferB-TDH were utilized as the samples. The samples were heated ranging from 35–90 °C that the increase of temperature speed was set to 90 °C per hour (Fig. S6C). The thermodynamic parameters were represented in Table S11.

Design of the artificial HNLs by GAOptimizer

Two of artificial *S*-hydroxynitrile lyase (HNL) were designed as follows. Structure of HNL from *Manihot esculenta* (MeHNL, PDB ID: 1DWO), which was energy-minimized by PyRosetta, and a library (M = 1) were used as the input data. Here, the library was prepared by downloading the homologs sequences of MeHNL by Blastp analysis with the condition that E-value and database were set to $1.0E^{-100}$ and non-redundant, respectively; total of 16 sequences could be obtained (Table S13). Utilizing the input data, the MeHNL was evolved computationally by GAOptimizer through 500 generations (N = 500); the HiSol score was adopted as fitness function (Fig. S8). The designed HNLs at 12^{th} and 500^{th} generation were adopted as research targets (Table S14). The sequence identity among the MeHNL and GaHNLs (Table S14) were shown in Table S15.

Overexpression and purification of the HNLs

A gene encoding two artificial HNLs (GaHNL-12gen and GaHNL-500gen, Table S14) was synthesized by GenScript and subcloned into pET28b vector at *Nde*I and *Xho*I site. The genes were optimized for the *E.coli* expression, and stop codon was added at the 3'-terminus of the genes. The plasmids coding MeHNL was obtained from the previous study [8]. The strain was cultivated in 1L LB browth containing 40 μg/mL kanamycin at 37 °C. The flask was kept on icy water for 30 min when the O.D.600 value was reached to the range from 0.6 to 0.8. After addition of 0.5 mM IPTG to the flask at final concentration, the cultivation was continued for 16 hr at 23 °C condition. The cells were suspended into bufferA-HNL (10 mM potassium phosphate [pH 7.0] and 50 mM NaCl). The suspended cells were sonicated, and the supernatant was collected by centrifugation at 22,000 g for 40 min. The supernatant was applied to HisTrap-HP column equilibrated by bufferA-HNL. After the column was washed with bufferA-HNL containing 70 mM imidazole, the HNL samples were eluted by bufferA-HNL containing 300 mM imidazole. The HNL samples were concentrated to 500 μL utilizing centrifugal filter unit, Amicon Ultra-15 (Merck Millipore). The samples were applied to Superdex 200 Increase 10/300 GL column (Cytiva, MA, USA) equilibrated by bufferA-HNL. The purity of the samples was confirmed by SDS-PAGE. The concentration of the samples was estimated by measuring absorption at 280 nm by NanoDrop One (Thermofisher Scientific, USA). Protein expression level of HNLs was estimated by measuring the amount of the samples obtained after the purification by Ni-affinity chromatography (Table S16).

Analysis of enzymatic properties of the HNLs

Enzymatic activity of the HNLs was measured by quantifying the concentration of the production of benzaldehyde with color developed method. All of the following experiments were performed in three replicates. The following assay buffer was prepared: 100 mM sodium citrate (pH 5.5), and 0.5–10 mM *rac*-mandelonitrile. The reaction was started by adding the purified HNL samples into the assay buffer which were dispensed into the cuvette. The initial velocity of the HNLs was estimated by monitoring time-dependent absorption change at 280 nm derived from benzaldehyde by UV-visible spectrometer (UV-2600i, Shimadzu); the molar extinction coefficient of benzaldehyde at 280 nm was 1376 $M^{-1}*cm^{-1}$. The enzyme kinetic parameters were estimated by fitting the initial velocity to the Michaelis-Menten equation with the non-linear least square method by ORIGIN software (Fig. S9A). The calculated parameters were represented in Table S17.

Thermal stability of the HNLs was analyzed by PEAQ-DSC. Total 22–196 μM of the purified HNLs in the bufferA-HNL were utilized as the samples. The samples were heated ranging from 50–100 °C that the increase of temperature speed was set to 90 °C per hour (Fig. S9B). The thermodynamic parameters obtained by fitting the DSC isotherms were represented in Table S18.

Crystallization and X-ray data collection of GaHNL-12gen

For the crystallization of the GaHNL-12gen, the samples were concentrated to more than 20 mg/mL. The samples were mixed with reservoir solution (20% (w/v) Polyethylene glycol 3350 and 0.2 M potassium chloride) as following molar ratio: 2:1 = protein:reservoir. The crystals of GaHNL-12gen appeared under the condition at 22 ºC. The obtained crystals were soaked into a cryo-reservoir composed of crystallization reservoirs containing 20% (v/v) glycerol. The soaked crystals were mounted and flash-cooled under a nitrogen stream at -173 ºC. Diffraction data were collected using PILATUS3 S2M at BL5A of Photon Factory (Tsukuba, Japan). The data were collected, scaled, and integrated by XDS[9]. The initial phase was determined by molecular replacement method with PHASER implemented in PHENIX[10] utilizing the structure of the crystal structure of MeHNL (PDB ID: 1DWP) as a template. Model building and structure refinement were performed by COOT[11] and REFMAC[12], respectively. All figures were prepared by PyMOL[13]. The crystallographic table was written in Table S19.

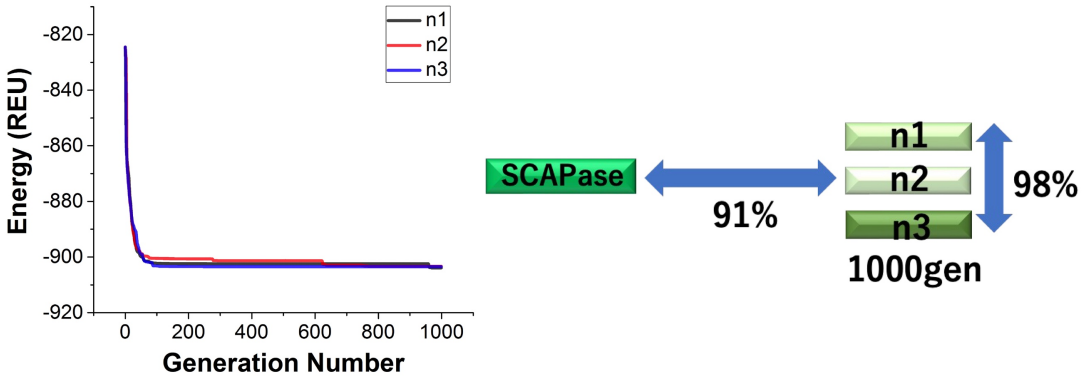

Fig. S1. Three independent run of GAOptimizer to minimize rosetta energy unit (REU) of native alkaline phosphatase (ALP), SCAPase. In each of the calculations (from n1 to n3), the REU values were converged to about -905 during 1000 generation run. The ALPs designed by GAOptimizer (n1–n3) shared high sequence identity to each other (>98%).

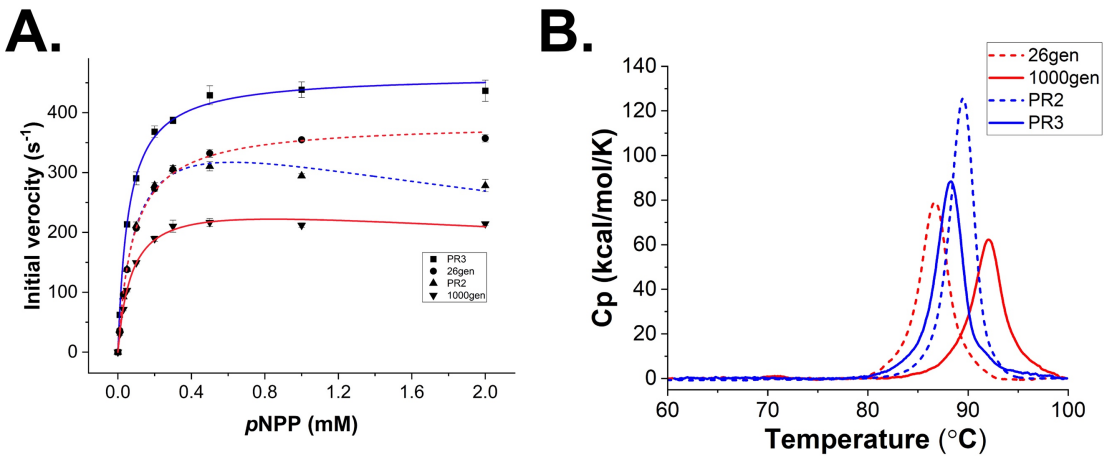

Fig. S2. Enzyme kinetic plots of the designed ALPs (A). In the figures, the data for the PROSS-2 (PR2), PROSS-3 (PR3), GaALP-26gen (26gen) and GaALP-1000gen (1000gen) were depicted as blue dotted line, blue straight line, red dotted line and red straight line, respectively. **DSC isotherms of the designed ALPs (B).** Small amount of the purified ALPs were obtained, so only the T_m values could be estimated. All of the measurements were performed by three replicates, and the data were shown as the mean ± S.D.

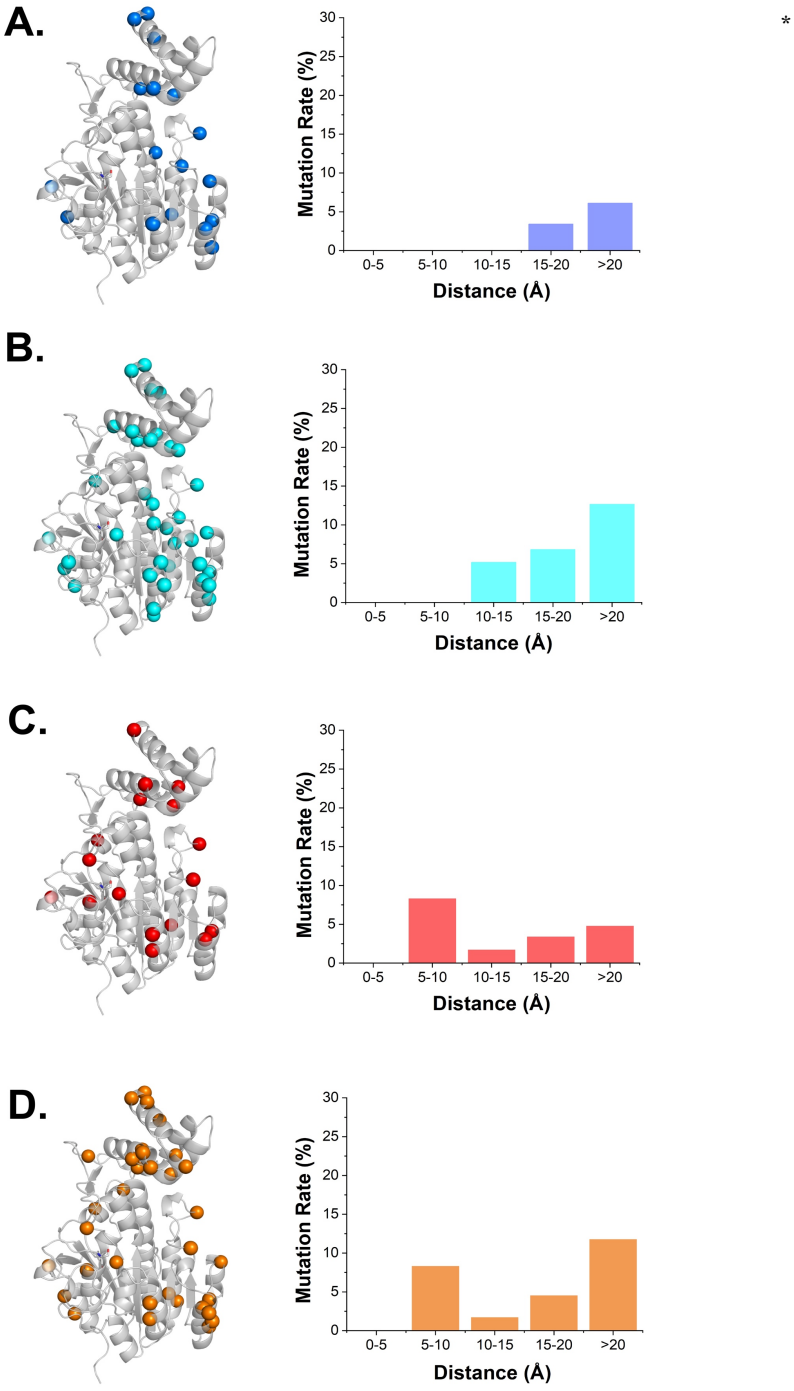

Fig. S3. Distribution of mutations sites and mutation rate at the five of the categorized regions by the distance from the catalytic residue of an oxygen atom of S58 between SCAPase and PROSS-2 (A), between SCAPase and PROSS-3 (B), between SCAPase and GaALP-26gen (C), and between SCAPase and GaALP-1000gen (D). The rate between CnTDH and GaTDH-26gen, and between CnTDH and GaTDH-2000gen were represented in B and D, respectively.

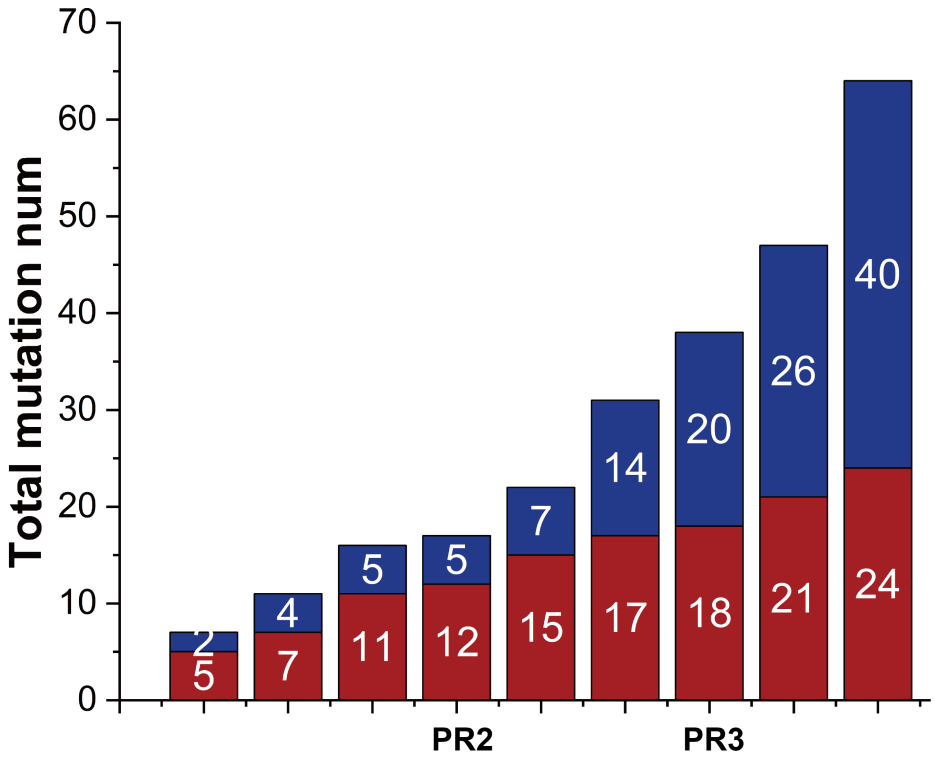

Fig. S4. Analysis of mutation concordance between GaALP-1000gen (1000gen) and the ALPs designed by PROSS: RROSS-2 (PR2) and PROSS-3 (PR3). The number of mutations were represented as stacked bar chart; the number of mutations which are commonly confirmed between the designed ALPs and 1000gen were represented as red bar, and the number of mutations which are uniquely confirmed in the designed ALPs were shown as blue bar, respectively.

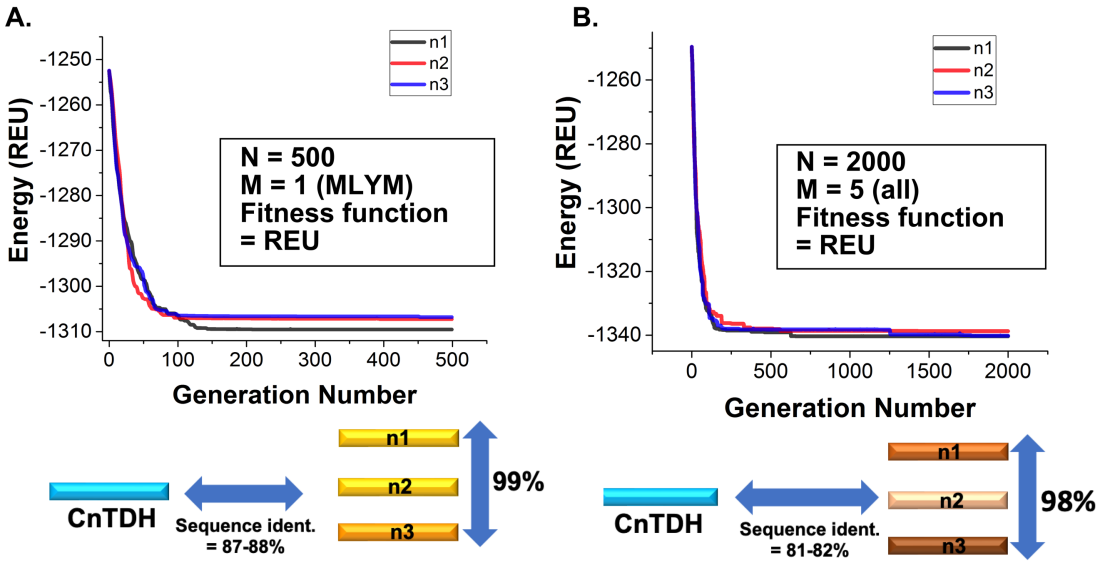

Fig. S5, Three independent run of GAOptimizer to minimize REU values of TDH from *Cupriavidus necator* (CnTDH) utilizing one (A) or five library data (B). In each of the calculations, the REU values were converged to about -1308 (A) and -1340 (B), respectively. The generated sequences by GAOptimizer (n1–n3) had high sequence identity to each other (>98%).

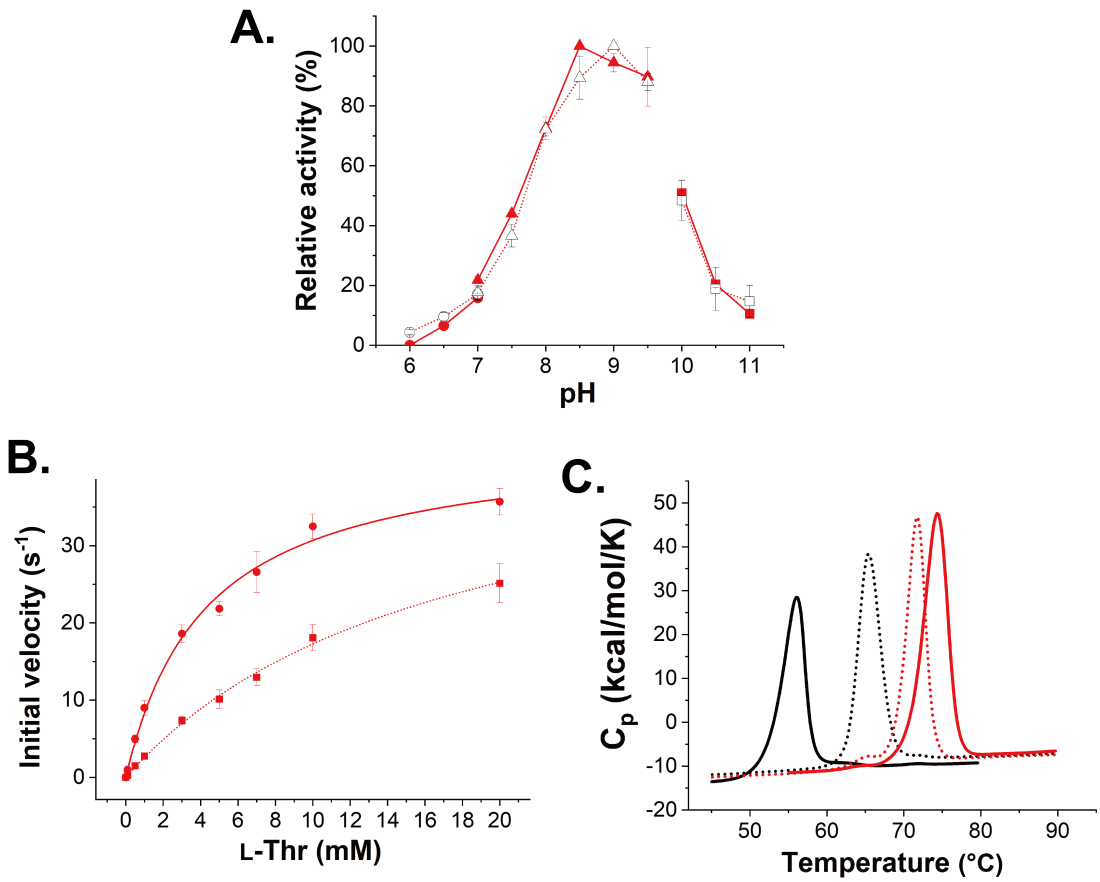

Fig. S6, Estimation for optimal pH values of GaTDH-26gen and GaTDH-2000gen (A). In the figures, the data of GaTDH-26gen and GaTDH-2000gen were represented as red straight line and red dotted line, respectively. The optimal pH values of GaTDH-26gen and GaTDH-2000gen were 8.5 and 9.0. **Enzyme kinetic plots of GaTDH-26gen and GaTDH-2000gen (B). DSC isotherms of CnTDH, FcTDH, GaTDH-26gen and GaTDH-2000gen (C).** The data for CnTDH and FcTDH were represented as black straight line and black dotted line, respectively. The thermodynamic parameters were indicated in Table S11. All of the measurements were performed by three replicates, and the data were shown as the mean ± S.D..

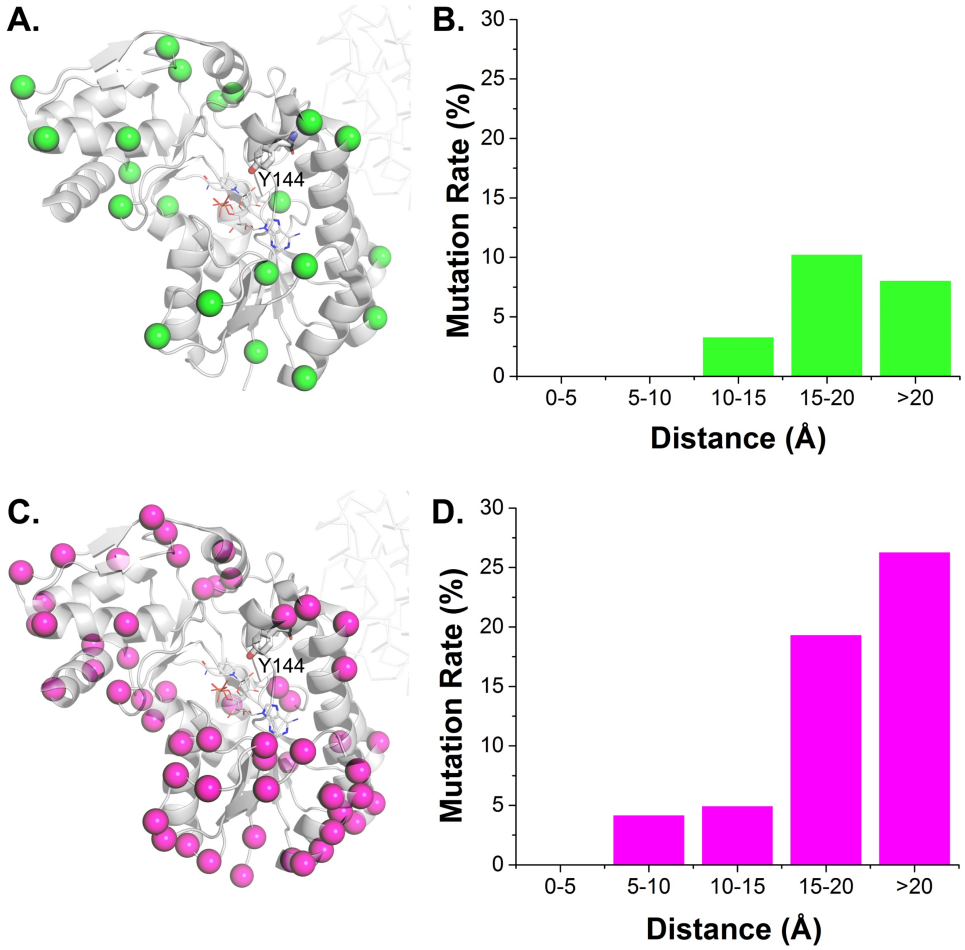

Fig. S7. Distribution of mutations sites between CnTDH and GaTDH-26gen (A), and between CnTDH and GaTDH-2000gen (C). Mutation rate at the five of the categorized regions by the distance from the catalytic residue of an oxygen atom of Y144 in CnTDH (B and D). The rate between CnTDH and GaTDH-26gen, and between CnTDH and GaTDH-2000gen were represented in B and D, respectively.

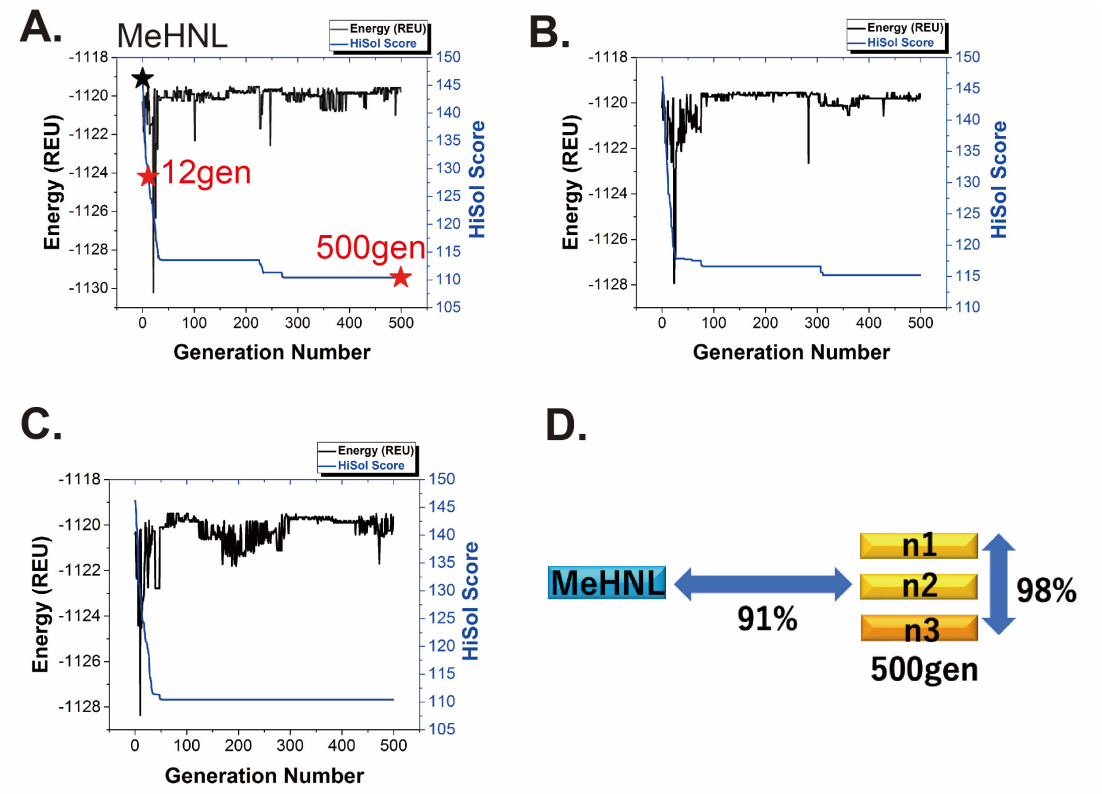

Fig. S8. Three independent run of GAOptimizer to minimize HiSol scores of MeHNL (A to C). The generation dependent changes of REU and HiSol scores were shown by black and blue lines, respectively. In the process, we set the generation number to 500. **The comparsion among the MeHNL and the generated three of HNLs by GAOptimizer (n1 to n3) (D).** As shown in the figure, the n1 to n3 shared high sequence identity (>98%) to each other.

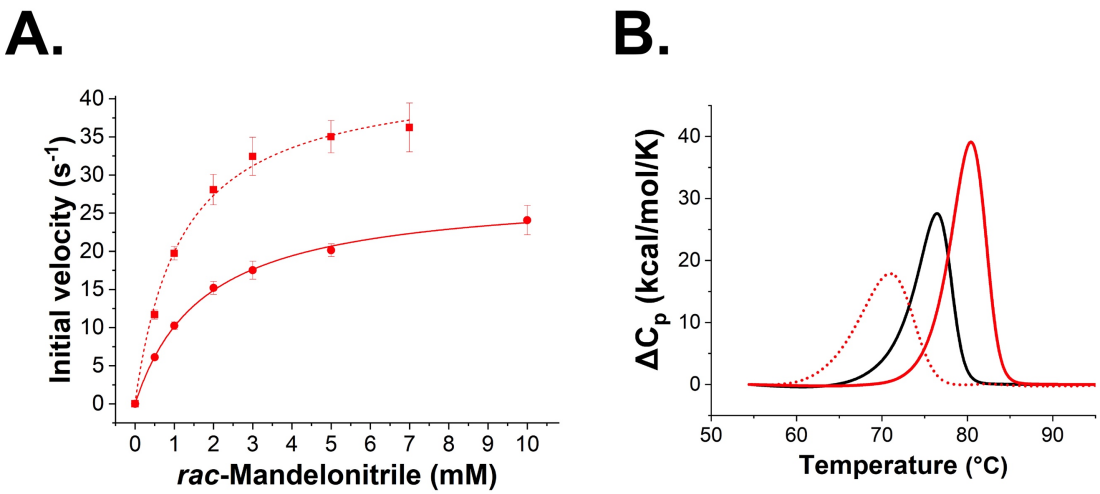

Fig. S9. Enzyme kinetic plots of GaHNLs (A). In the figures, the data for the GaHNL-12gen and GaHNL-500gen were depicted as red straight line and red dotted line, respectively. **DSC isotherms of MeHNL and GaHNLs (B).** The data for MeHNL was shown by black line. All of the measurements were performed by three replicates, and the data were shown as the mean ± S.D.

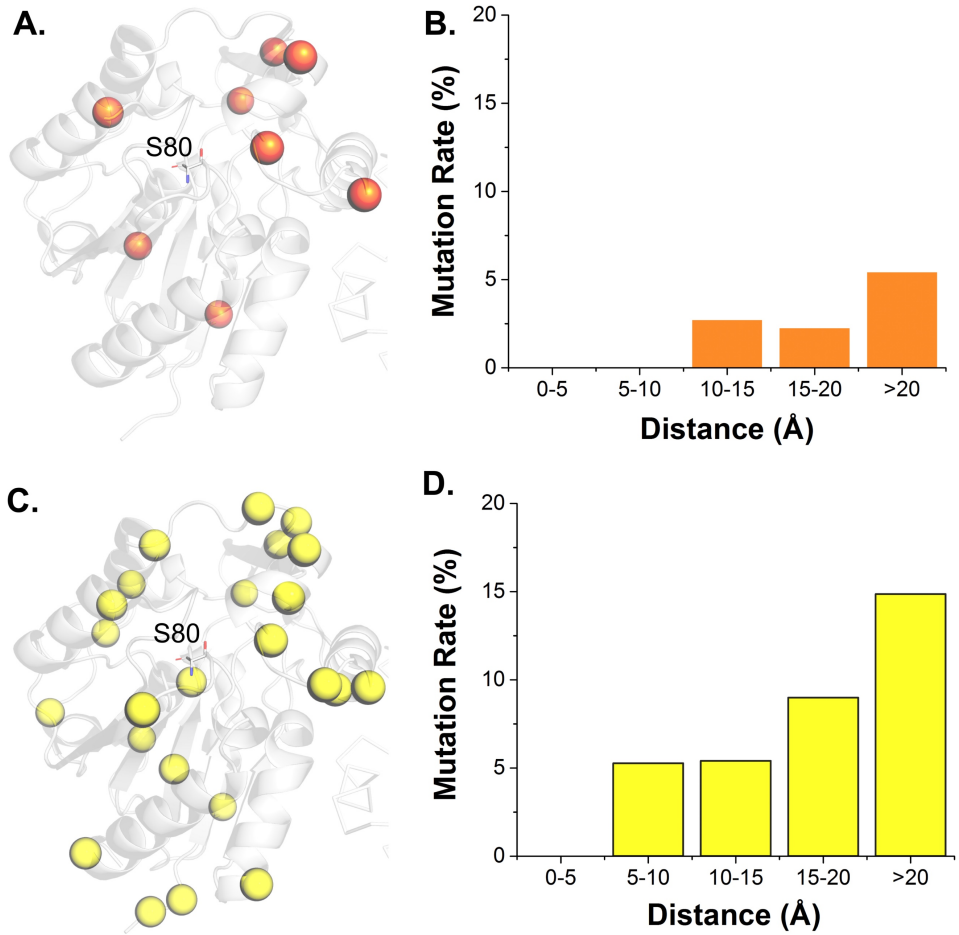

Fig. S10. Distribution of mutation sites between MeHNL and GaHNL-12gen (A), and between MeHNL and GaHNL-500gen (C). Mutation rate at the five of the categorized regions by distance from an oxygen atom of the catalytic residue S80 (B and D). The rate between MeHNL and GaHNL-12gen (B), and between MeHNL and GaHNL-500gen (D) were represented by orange and yellow, respectively.

Table S1. The contents of 166 homologs sequences (curated-lib) of SCAPase that were utilized to design GaALPs and PROSS-2 and PROSS-3.

Accession
3A52_A, WP_202287832.1, GIU03776.1, WP_020914455.1, GIU25515.1, GIU50552.1, GIU06639.1, WP_012275828.1, WP_012153864.1, WP_202720781.1, WP_028765620.1, WP_083698135.1, WP_102436773.1, WP_112354238.1, WP_182698386.1, WP_144214525.1, GIU52220.1, WP_119977275.1, WP_124729648.1, KPZ73564.1, WP_220069120.1, WP_162022255.1, WP_188840215.1, WP_123778009.1, WP_137223167.1, WP_011788159.1, WP_167679935.1, WP_101032902.1, PMG79422.1, NRB25372.1, WP_188957235.1, WP_220050200.1, WP_142872910.1, WP_143562653.1, AZQ09258.1, WP_220079110.1, WP_208162623.1, WP_203325957.1, WP_220043357.1, MBL4817445.1, WP_011758691.1, WP_220081766.1, WP_088905463.1, WP_188738312.1, WP_152829346.1, WP_012144308.1, WP_041509999.1, WP_126519334.1, WP_188918770.1, WP_220059077.1, WP_160054992.1, WP_011867022.1, WP_076414270.1, WP_188926203.1, WP_179948374.1, WP_012326721.1, WP_179023410.1, WP_220062097.1, WP_144204333.1, WP_126506181.1, WP_207321869.1, WP_083779945.1, WP_133037816.1, WP_207380926.1, WP_220108280.1, WP_160793664.1, WP_100142640.1, WP_167868452.1, WP_153664735.1, WP_207354342.1, WP_213681011.1, WP_011497402.1, WP_182677889.1, WP_115137321.1, WP_218528304.1, WP_188919395.1, WP_121853693.1, WP_205000457.1, MBE8167950.1, WP_129948572.1, WP_133405716.1, WP_095505577.1, WP_140933736.1, WP_095499917.1, NRA83621.1, WP_199436884.1, WP_138611305.1, WP_045395624.1, WP_138630223.1, WP_055013637.1, WP_038886767.1, WP_047477531.1, WP_138560387.1, MAD90128.1, ARR08908.1, WP_130254814.1, WP_216056473.1, WP_005531546.1, PKF56028.1, WP_105051004.1, WP_138566559.1, WP_025790956.1, WP_114092301.1, WP_055199013.1, WP_180311877.1, WP_138479397.1, WP_064664952.1, WP_166376628.1, WP_025533039.1, WP_215932660.1, WP_197111533.1, WP_210798434.1, WP_140956458.1, WP_193156703.1, WP_039988659.1, HDY90941.1, WP_023585058.1, WP_104027151.1, ROQ28701.1, WP_010556198.1, WP_058559349.1, CCQ10631.1, WP_105254674.1, MBR9790916.1, HEA17896.1, WP_123324458.1, WP_110429021.1, WP_071815446.1, OFI34093.1, NVK54054.1, WP_100656721.1, WP_212547423.1, WP_208844773.1, WP_124028333.1, WP_218312901.1, MBT4883918.1, WP_216030145.1, WP_076010958.1, HAQ49892.1, BBO28745.1, AGP78689.1, WP_008900589.1, RUM31296.1, WP_108565950.1, WP_218352591.1, MBU1619430.1, MBP8227796.1, WP_170308427.1, WP_184424094.1, WP_008484183.1, WP_083638607.1, MAQ02290.1, WP_083638571.1, HCB09564.1, WP_214996269.1, WP_039218212.1, WP_118493197.1, WP_179984084.1, WP_202942194.1, WP_115591509.1, GFD96185.1, WP_198816002.1, GFD90299.1, WP_141152458.1, WP_136782224.1, WP_119502039.1,

Table. S2. The summary for the information of the designed ALPs.

Sample name	Design tool	Library	Initial structure	number of Mutations
SCAPase (3A52.pdb)	-	-	-	0
PROSS-1 (PR1)	PROSS	auto	3A52.pdb	34
GaALP-26gen (26gen)	GAOptimizer	curated-lib	3A52-EM.pdb	17
GaALP-1000gen (1000gen)	GAOptimizer	curated-lib	3A52-EM.pdb	34
PROSS-2 (PR2)	PROSS	curated-lib	3A52-EM.pdb	17
PROSS-3 (PR3)	PROSS	curated-lib	3A52-EM.pdb	38

Table S3. Protein sequences of the designed ALPs

>GaALP-26gen
MGMIIMIGDGMGPAYTTAYRYFQDNPDTEEIEQTVFDRLLVGMASTYPARESGYVTDSA
AAATALATGFKSYNGAIAVDINKRPLTTIMQMAKARGMSTGVAVTAQVNHATPAAFLTHN
ESRKNYEEIAASMLKSDADVILGGGRKYFSEALLSQFSAKGYQYITELAQLDSITQPKVL
GLFADVQLPWVIDDKDANTLSKLTQKALDLLSQNEKGFVLLVEGSLIDWAGHNNDIATAM
AEMQGFANAIEVVEQYIRQHPDTLLVVTADHSTGGLSIGANGEYQWDTKLLKGIKASPAS
IATHAIAADDWQAGVNQLLGFDVNSEELQQLTNARMQGKSALEVALKKIIDTRSYTGWT
TSGHTGVDVQVFAMGPAADLFKGHQDNTHIAEKMMSLLPKVN

>GaALP-1000gen
MGMIIMIGDGMGPAYTTAYRYYQDNPDTEEIEQTVFDRLLVGMASTYPARESGYVTDSA
AAATALATGFKSYNGAIAVDINKRPLTTIMQMAKARGMSTGVAVTAQVNHATPAAFLTHN
ESRKNYEEIAASMLKSDADVILGGGRKYFSENLLAQFKAKGYQYITELNQLDSITQPKVL
GLFADVQLPWVIDDKDANTLSKLTQKALELLSQNEKGFVLLVEGSLIDWAGHNNDIATAM
AEMQGFANAIEVVEQYIRQHPDTLLVVTADHSTGGLSIGANGEYRWDTELLKGIKASPAS
IAKHAIAADDWQAGVNQLLGFEVNDEELAQLTNARMQGKKALETAIKKIIDTRSYTGWTT
SGHTGVDVQVFAMGPAADLFKGHQDNTDIAKKMMSLLPKVN

>PROSS-1
MGMIIMIGDGMGPAYTTAYRYFQDNPDTEEIEQTVFDRLLVGMASTYPARESGYVTDSA
AAATALATGFKTYNGAIAVDINKRPLTTIMEMAKARGMSTGVVVTAQVNHATPAAFLTHN
ESRKNYEEIAQDLLKSDADVILGGGRKYFSEALLSQFKAKGYQYITELAQLDSITQPKVL
GLFAEVQLPWVIDDTDANRLAKMTQKALDLLSQNPKGFVLLVEASLIDWAGHNNDIATA
MAEMQDFANAIEVVEQYIRQHPDTLLVVTADHNTGGLSIGANGEYQWDTKLLKNIKASP
ASIAEHAIAADDWQATVNQLLGFDLNDEELQQLTNARMQGKETLEVALKKIINTRSYTGW
TTSGHTGVDVQVFAMGPWADLFKGHQDNTDIAKKMMSLLPKVN

>PROSS-2
MGMIIMVGDGMGPAYTSAYRYFQDNPDTEEIEQTVFDRLLVGMASTYPARESGYVTDSA
ASATALATGFKSYNGAIAVDINKRPLTTIMQMAKARGMSTGVAVTAQVNHATPAAFLTHN
ESRKNYEEIAADMLKSDADVILGGGRKYFSEALLSQFKAKGYQYITEFAQLDSITQPKVL
GLFAEVQLPWVIDDKDANRLSKLTQKALDLLSQNEKGFVLLVEGSLIDWAGHNNDIATAM
AEMQDFANAIEVVEQYIRQHPDTLLVVTADHNTGGLSIGANGEYQWDTKLLKGIKASPAS
IATHAIAADDWQAGVNQQLGFDLNDEELQQLTNARMQGKETLEVALKKIIDTRSYTGWT
TSGHTGVDVQVFAMGPAADLFKGHQDNTDIAEKMMSLLPKVN

>PROSS-3
MGMIIMIGDGMGPAYTTAYRYFQDNPDTEEIEQTVFDRLLVGMASTYPARESGYVTDSA
ASATALATGFKTYNGAIAVDINKRPLTTIMQMAKARGMSTGVAVTSQVNHATPAAFLTHN
ESRKNYDEIAQDMLNSDADVILGGGQKYFPEELLSQFKAKGYQYITEFAQLDSITQPKVL
GLFAEVQLPWVIDDKDANKLAKMTQKALDLLSQNPKGFVLLVEGSLIDWAGHNNDIATA
MAEMHEFANAIEVVEQYIRQHPDTLLVVTADHNTGGLSIGANGEYQWDTELLKNIKASPA
TIATHAIAADDWQAGVNQQLGFELNDEELQQLTNARMQGKETLETALKKIIDTRTYTGWT
TSGHTGVDVQVFAMGPAADLFKGHQDNTDIAKKMMSLLPKVN

Table S4. Sequence identity between the native and the designed ALPs[a]

	SCAPase	GaALP-26gen	PROSS-2	GaALP-1000gen	PROSS-3	PROSS-1
SCAPase	100%	95.8%	95.8%	91.5%	90.5%	91.5%
GaALP-26gen	95.8%	100%	96.0%	95.8%	91.8%	93.3%
PROSS-2	95.8%	96.0%	100%	93.5%	94.3%	94.8%
GaALP-1000gen	91.5%	95.8%	93.5%	100%	91.5%	91.5%
PROSS-3	90.5%	91.8%	94.3%	91.5%	100%	93.5%
PROSS-1	91.5%	93.3%	94.8%	91.5%	93.5%	100%

[a]The sequence identity was cited from the output of MAFFT titled “Percent identity matrix”[2].

Table S5. Protein expression level of ALPs by *E.coli* expression system.

Sample[a]	Total protein (mg/L)	Specific activity (U/mg)
SCAPase (template)	n.d.	About 900[b]
PROSS-1	81.3 ± 25.3	65.9
GaALP-26gen	17.0 ± 2.3	495.9
GaALP-1000gen	21.5 ± 3.9	303.2
PROSS-2	32.5 ± 7.2	427.7
PROSS-3	26.8 ± 14.5	598.6

[a]The samples were expressed by adding His-tag at C-terminal region; amount of the total proteins were estimated by quantifying the amount of the samples after the purification by Ni-affinity chromatography.

[b]The parameter (at 25 °C) was estimated from the figure 3B in the reference [3]. The parameter was obtained by analyzing the specific activity which was expressed as GST fusion protein.

Table S6. Enzyme kinetic parameters of the designed ALPs.

	k_{cat}	K_m	k_{cat}/K_m	T_m[a]
Sample name	s^{-1}	mM	$s^{-1}\,mM^{-1}$	°C
SCAPase	n.d.	n.d.	n.d.	n.d.
PROSS-1	57.5 ± 1.7	0.054 ± 0.005	1060	>100
GaALP-26gen	383.2 ± 4.1	0.085 ± 0.004	4510	86.6 ± 0.1
GaALP-1000gen	261.7 ± 8.0	0.075 ± 0.006	3490	92.2 ± 0.2
PROSS-2	414.0 ± 17.5	0.095 ± 0.010	4360	89.3 ± 0.1
PROSS-3	463.4 ± 5.8	0.058 ± 0.004	7990	89.1 ± 0.3

[a]The T_m values were estimated by DSC analysis.

Table S7. Protein sequences of the designed TDHs generated by GAOptimizer

>GaTDH-26gen
MGKPRILIIGANGQIGSELALALAERYGRTNVITSDVVPEGRHEHLTHEMLDATDRGELAT
VVEKHGITQVYLLAAALSATGEKNPLWAWNLNMTSLLNVLELARETGLERVFWPSSIAAF
GPTTPKEQTPQKTVMEPTTVYGISKQAGEGWCRWYHAKHGVDVRSVRYPGLISYKTP
PGGGTTDYAVEIFHAAVKGEPYTCFLAEDTALPMMYMPDAIRATIELMEAPADKLSERGA
YNIAGMSFTPAQIAAAIREHVPDFQIRYEPDYRQAIAQGWPDSIDDSVARADWGWKPQY
DLKEMVADMLANLK

>GaTDH-2000gen
MEKPRILIIGACGQIGTELALALAERYGNENVITSDIVPEGLHEKLTHEYLDATDKKALEEV
VKKHNITQVYLLAAALSATAEKNPLWAWKLNMTSLLNVLELARELGLERVFWPSSIAAFG
PTTPKDQTPQKTVMEPTTVYGISKQAGEGWCRWYFAKHGVDVRSVRYPGLISYKTPPG
GGTTDYAVEIFHAAVKGEPYTCFLAEDTALPMMYMDDAIKATIELMEAPASKLSIRGAYNI
AGMSFTPKQIAAAIRKHVPDFEIRYKPDYRQAIADGWPNSIDDSVARKDWGWKPEYDLD
KMVADMLKNLK

Table S8. Sequence identity between the native and the designed TDHs

	CnTDH	GaTDH-26gen	GaTDH-2000gen	FcTDH
CnTDH	100%	92.9%	81.6%	67.0%
GaTDH-26gen	92.9%	100%	88.4%	71.3%
GaTDH-2000gen	81.6%	88.4%	100%	71.6%
[a]FcTDH	67.0%	71.3%	76.1%	100%

[a]The sequence of FcTDH was cited from the reference [6]. The FcTDH is corresponding with the sequence of FcTDH-VLYM.

Table S9. Protein expression level of TDHs by *E.coli* expression system.

Sample Name	mg/L
CnTDH[a]	87.7 ± 2.9
GaTDH-26gen	172.0 ± 14.8
GaTDH-2000gen	138.7 ± 17.7
FcTDH	149.4 ± 16.9

Table S10. Enzyme kinetic parameters of TDHs at optimal pH values[c].

	k_{cat}	K_m	k_{cat}/K_m
Sample name	s^{-1}	mM	s^{-1} mM^{-1}
CnTDH[a]	118.9 ± 6.0	17.4 ± 1.6	6.8 ± 0.7
GaTDH-26gen	43.7 ± 1.9	4.3 ± 0.5	10.3 ± 1.3
GaTDH-2000gen	47.0 ± 3.0	17.2 ± 1.8	2.7 ± 0.3
FcTDH[b]	8.2 ± 0.2	0.80 ± 0.1	10.2 ± 1.3

[a]The parameters were cited from the reference [7].
[b]The parameters were cited from the reference [6].
[c]Activity of GaTDH-26gen and GaTDH-2000gen was measured at pH 8.5 and 9.0 condition, respectively.

Table S11. Thermodynamic parameters of TDHs estimated by DSC analysis

	T_m (°C)	ΔH_{cal} (kcal/mol)	ΔH_{VH} (kcal/mol)
CnTDH	55.7 ± 0.0	180 ± 0.9	237 ± 1.4
GaTDH-26gen	74.1 ± 0.1	232 ± 0.5	212 ± 2.2
GaTDH-2000gen	71.6 ± 0.0	221 ± 1.2	295 ± 1.9
FcTDH	65.5 ± 0.0	206 ± 0.8	250 ± 0.5

Table S12. Improvement of thermostability and REU values per one mutations among the designed TDHs

	ΔT_m (°C)[a]	Number of mutations	Increase of ΔT_m per one mutation[b]	Increase of *ΔREU* per one mutation[b]
GaTDH-26gen	18.4	23	0.801	2.04
GaTDH-2000gen	15.9	56	0.284	1.63
FcTDH	9.9	107	0.092	0 (n.d.)
CnTDH	0 (n.d.)	0 (n.d.)	0 (n.d.)	0 (n.d.)

[a]The ΔT_m value was calculated by subtracting the T_m values of the designed TDHs from the T_m value of CnTDH.
[b]The parameters were calculated by dividing the values (ΔT_m or *ΔREU*) from the number of mutations.

Table S13. The library contents of 16 homologs sequences of MeHNL that were utilized to design GaHNLs[a].

Accession	Definition	Species
1DWO_A	HYDROXYNITRILE LYASE	*Manihot esculenta*
XP_021633434.1	(*S*)-hydroxynitrile lyase-like	*Manihot esculenta*
XP_021684455.1	(*S*)-hydroxynitrile lyase-like	*Hevea brasiliensis*
5TDX_A	Ancestral Hydroxynitrile Lyase 1	Artificial sequence
XP_021647581.1	(*S*)-hydroxynitrile lyase	*Hevea brasiliensis*
XP_043805401.1	(*S*)-hydroxynitrile lyase	*Manihot esculenta*
XP_021637831.1	(*S*)-hydroxynitrile lyase-like	*Hevea brasiliensis*
XP_021631609.1	(*S*)-hydroxynitrile lyase	*Manihot esculenta*
XP_021633808.1	(*S*)-hydroxynitrile lyase	*Manihot esculenta*
KAF2292990.1	hypothetical protein GH714_034648	*Hevea brasiliensis*
KAF2284003.1	hypothetical protein GH714_017898	*Hevea brasiliensis*
XP_021631232.1	(*S*)-hydroxynitrile lyase	*Manihot esculenta*
KAF2308129.1	hypothetical protein GH714_035553	*Hevea brasiliensis*
XP_021648645.1	(*S*)-hydroxynitrile lyase-like	*Hevea brasiliensis*
XP_021631480.1	(*S*)-hydroxynitrile lyase	*Manihot esculenta*
XP_021659844.1	(*S*)-hydroxynitrile lyase-like	*Hevea brasiliensis*

[a]The sequence identity among the library was distributed in the range from 62 to 89%. The library was prepared by submitting the sequence of MeHNL to Blastp by the following condition: the E-value was set to $1.0E^{-90}$.

Table S14. Protein sequences of the designed HNLs generated by GAOptimizer

>GaHNL-12gen
MVTAHFVLIHTICHGAWIWHKLKPALERAGHKVTALDMAASGIDPRQIEQINSFDEYSEPL
LTFLESLPQGEKVIIVGESCAGLNIAIAADKYVDKIAAGVFHNSLLPDTVHSPSYVVEKLLE
SFPDWRDTEYFTYTNNTGETITTMKLGFKLLRENLYTKCTDEEYELAKMVMRKGSLFQN
VLAQRPKFTEKGYGSIKKVYIWTDQDKIFLPDFQRWQIANYKPDKVYQVQGGDHKLQLT
KTEEVAHILQEVADAYA

>GaHNL-500gen
MAVAHFVLIHTICHGAWIWHKLKPALEAAGHKVTALDMAASGIDPRQIEQINSFDEYSEPL
LTFLESLPQGEKVILVGESCGGLNIAIAADKYVDKIAAGVFHNSLLPDTVHSPSYVVEKLM
ESFPDWKDTVYFTYTNNTGETITTMKLGFKLMRENLYTKCPDEDYELAKMVMRKGSLF
QNVLAKREKFTKKGYGSIKKVYIWTDQDKIFLPDFQRWQIANYKPDKVYQVPGGDHKLQ
LTKTEEVAHILQEVADTYA

Table S15. Sequence identity between the native and the designed HNLs

	MeHNL	GaHNL-12gen	GaHNL-500gen
MeHNL	100%	96.9%	90.7%
GaHNL-12gen	96.9%	100%	93.8%
GaHNL-500gen	90.7%	93.8%	100%

Table S16. Protein expression level of HNLs by *E.coli* expression system.

	mg/L
MeHNL	25.1 ± 1.8
GaHNL-12gen	194.2 ± 20.9
GaHNL-500gen	22.9 ± 0.9

Table S17. Enzyme kinetic parameters of the native and the designed HNLs.

	k_{cat}	K_m	k_{cat}/K_m
Sample name	s^{-1}	mM	s^{-1} mM^{-1}
MeHNL[a]	34.5	5.2	6.7
GaHNL-12gen	31.0 ± 0.3	1.7 ± 0.1	18.2 ± 1.1
GaHNL-500gen	47.0 ± 1.0	1.3 ± 0.1	36.2 ± 2.9

[a]The parameters were cited from the reference [8].

Table S18. Thermodynamic parameters of HNLs estimated by DSC analysis

	T_m	ΔH_{cal}	ΔH_{vh}
Sample name	°C	kcal/mol	kcal/mol
MeHNL	75.4 ± 0.5	130.7 ± 23.3	156.8 ± 14.2
GaHNL-12gen	80.0 ± 0.0	213.7 ± 6.0	183.5 ± 2.6
GaHNL-500gen	70.5 ± 0.1	145.8 ± 1.9	120.3 ± 3.1

Table S19. Statistics of X-ray diffraction data collection of GaHNL-12gen ligand free form.

	GaHNL-12gen
Space group	$P2_12_12_1$
Unit cell parameters	
a (Å)	85.7
b (Å)	86.4
c (Å)	135.0
α (degree)	90.0
β (degree)	90.0
γ (degree)	90.0
X-ray source	PF BL-5A
Wavelength (Å)	1.00
Resolution (Å)	45.2-1.69 (1.78-1.69)
No. of reflections[a]	2210859
No. of unique reflections	112048
Completeness (%)	100.0 (99.7)
I/sig (I)	31.8 (6.8)
R_{merge}[b]	0.072 (0.470)
$CC_{1/2}$	1.00 (0.979)
B of Wilson plot (Å^2)	12.3
R[c]	0.160
R_{free}[d]	0.198
RMSD of geometry	
Bond length (Å)	0.006
Bond angles (deg)	0.892
Geometry	
Ramachandran outlier (%)	0.0
Ramachandran favored (%)	100
PDB entry	8J82

[a.] Sigma cutoff was set to none (F > 0σF).

[b.] $R_{merge} = \Sigma_h\Sigma_i|I_i(h)-\langle I(h)\rangle|/\ \Sigma_h\ I(h)$, where $I_i(h)$ is the i^{th} measurement of reflection h, and $\langle I(h)\rangle$ is the mean value of the symmetry-related reflection intensities. Values in brackets are for the shell of the highest resolution.

[c.] $R = \Sigma||F_o|-|F_c\ ||/\ \Sigma|F_o\ |$, where F_o and F_c are the observed and calculated structure factors used in the refinement, respectively.

[d.] R_{free} is the *R*-factor calculated using 5% of the reflections chosen at random and omitted from the refinement.

References

[1] Chaudhury, S., Lyskov, S., and Gray, J. J. (2010) PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta, *Bioinformatics 26*, 689-691.

[2] Yamada, K. D., Tomii, K., and Katoh, K. (2016) Application of the MAFFT sequence alignment program to large data-reexamination of the usefulness of chained guide trees, *Bioinformatics 32*, 3246-3251.

[3] Murakawa, T., Yamagata, H., Tsuruta, H., and Aizono, Y. (2002) Cloning of cold-active alkaline phosphatase gene of a psychrophile, Shewanella sp., and expression of the recombinant enzyme, *Biosci Biotechnol Biochem 66*, 754-761.

[4] Peleg, Y., Vincentelli, R., Collins, B. M., Chen, K. E., Livingstone, E. K., Weeratunga, S., Leneva, N., Guo, Q., Remans, K., Perez, K., Bjerga, G. E. K., Larsen, O., Vanek, O., Skorepa, O., Jacquemin, S., Poterszman, A., Kjaer, S., Christodoulou, E., Albeck, S., Dym, O., Ainbinder, E., Unger, T., Schuetz, A., Matthes, S., Bader, M., de Marco, A., Storici, P., Semrau, M. S., Stolt-Bergner, P., Aigner, C., Suppmann, S., Goldenzweig, A., and Fleishman, S. J. (2021) Community-Wide Experimental Evaluation of the PROSS Stability-Design Method, *J Mol Biol 433*, 166964.

[5] Goldenzweig, A., Goldsmith, M., Hill, S. E., Gertman, O., Laurino, P., Ashani, Y., Dym, O., Unger, T., Albeck, S., Prilusky, J., Lieberman, R. L., Aharoni, A., Silman, I., Sussman, J. L., Tawfik, D. S., and Fleishman, S. J. (2016) Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability, *Mol Cell 63*, 337-346.

[6] Motoyama, T., Hiramatsu, N., Asano, Y., Nakano, S., and Ito, S. (2020) Protein Sequence Selection Method That Enables Full Consensus Design of Artificial l-Threonine 3-Dehydrogenases with Unique Enzymatic Properties, *Biochemistry 59*, 3823-3833.

[7] Nakano, S., Motoyama, T., Miyashita, Y., Ishizuka, Y., Matsuo, N., Tokiwa, H., Shinoda, S., Asano, Y., and Ito, S. (2018) Benchmark Analysis of Native and Artificial NAD(+)-Dependent Enzymes Generated by a Sequence-Based Design Method with or without Phylogenetic Data, *Biochemistry 57*, 3722-3732.

[8] Asano, Y., Dadashipour, M., Yamazaki, M., Doi, N., and Komeda, H. (2011) Functional expression of a plant hydroxynitrile lyase in Escherichia coli by directed evolution: creation and characterization of highly in vivo soluble mutants, *Protein Engineering Design & Selection 24*, 607-616.

[9] Kabsch, W. (2010) Xds, *Acta Crystallogr D Biol Crystallogr 66*, 125-132.

[10] Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner,

R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr D Biol Crystallogr 66*, 213-221.

[11] Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and Development of Coot, *Acta Crystallogr. D Biol. Crystallogr. 66*, 486—501.

[12] Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the Refinement of Macromolecular Crystal Structures, *Acta Crystallogr. D Biol. Crystallogr. 67*, 355—367.

[13] Delano, W., L. (2002) The PyMOL Molecular Graphics System, (Delano Scientific, S. C., CA., Ed.).